Engineering

Industrial & Management Engineering fields

Okayama University *Year* 2001

Adaptive state construction for reinforcement learning and its application to robot navigation problems

Hisashi Handa* Akira Ninomiya† Tadashi Horiuchi‡
Tadataka Konishi** Mitsuru Baba††

*Okayama University
†Okayama University
‡Matsue National College of Technology
**Chugoku Polytechnic College
††Okayama University

ADAPTIVE STATE CONSTRUCTION FOR REINFORCEMENT LEARNING AND ITS APPLICATION TO ROBOT NAVIGATION PROBLEMS

HISASHI HANDA*, AKIRA NINOMIYA*, TADASHI HORIUCHI, TADATAKA KONISHI***, and MITSURU BABA ***

*Faculty of Engineering, Okayama University, Okayama, Japan 700-8530
**Dept. of Information Engineering, Matsue National College of Technology, Matsue, Japan 690-8518
***Chugoku Polytechnic College, Okayama, Japan 710-0251

Abstract

This paper applies our state construction method by ART Neural Network to Robot Navigation Problems. Agents in this paper consist of ART Neural Network and Contradiction Resolution Mechanism. The ART Neural Network serves as a mean of state recognition which maps stimulus inputs to a certain state and state construction which creates a new state when a current stimulus input cannot be categorized into any known states. On the other hand, the Contradiction Resolution Mechanism (CRM) uses agents' state transition table to detect inconsistency among constructed states. In the proposed method, two kinds of inconsistency for the CRM are introduced: "Different results caused by the same states and the same actions" and "Contradiction due to ambiguous states." The simulation results on the robot navigation problems confirm us the effectiveness of the proposed method.

Keywords

Adaptive State Construction, ART Neural Network, Reinforcement Learning

1 Introduction

Recently, we have proposed a novel incremental state construction method for Reinforcement Learning Agents which consists of ART Neural Network and Contradiction Resolution Mechanism [1]. The ART Neural Network serves as a mean of state recognition which maps stimulus inputs to a certain state and state construction which creates a new state when a current stimulus input cannot be categorized into any known states. On the other hand, the Contradiction Resolution Mechanism (CRM) uses agents' state transition table to detect inconsistency among constructed states. The state transition table is constituted by past state transition tuple which consists of a recognized state at certain time, an action at that time, and a state recognized at the next time. In the proposed method, two kinds of inconsistency for the CRM are introduced: "Different results caused by the same states and the same actions" and "Contradiction due to ambiguous states." In the CRM, the first inconsistent situation is resolved by creating a new state whose weights of ART Neural Network are the same as stimulus percepted at that time. The CRM maintains bias, which decides recognized states for overlapped states, in order to resolve the second inconsistent situation.

In this paper, we adopt the proposed method to a robot navigation problem. The simulation results confirm us the effectiveness of the proposed method: Especially, (1) the proposed method can autonomously constitute a map from stimulus to states based upon agent's experience. That is, it can reduce the effort of designers of Reinforcement Learning Agents (2) Unlike tile coding used for conventional reinforcement learning methods, such as Q-Learning, SALSA, and so on [2]-[4], it requires fewer elements in the Q-Table, i.e., less computational memory.

2 Related Works

Incremental state construction methods have been studied by many researchers [5]-[10]. Many of them used reinforcement signals to delineate states more precisely. In the proposed method, we adopt state transition information to acquire adequate state construction. Dubrawski and Reignier proposed perceptual state categorization method using Fuzzy-ART Neural Network [5]. Our method utilizes modified ART Neural Network based on distance between input vectors and refines state constitution by using the notion of contradiction. The notion of contradiction used in this work is inspired by Piaget's one [11]. In his work, the notion of contradiction

0-7803-7087-2/01/$10.00 © 2001 IEEE

is classified into three categories from the observation of children' behavior:

1. Contradictions such that it looks that the same actions yield the different results.
2. Contradictions characterized by incomplete disagreement among certain classes
3. Contradictions caused by incorrect reasoning, especially incorrect implication.

In his work, he concluded that contradictions are emerged from inconsistent complements. Two kinds of contradictions introduced in this paper, i.e., "Different results caused by the same states and the same actions" and "Contradiction due to ambiguous states," are belonging to category 1. and 2., respectively.

3 The Proposed Method

The diagram of our approach is depicted in Fig. 1. As depicted in this figure, we assume continuous inputs from environments, such like sensors, cameras and so on. For the sake of using traditional reinforcement algorithms, discrete state of agents is decided from the continuous inputs. In this paper, we adopt a kind of Adaptive Resonance Theory (ART) originally proposed by Grossberg as a map from such continuous inputs to discrete states [12]. Then, agents carry out proper action associated to such state based on this action selection mechanism, and recognize new perceptual inputs from the environments again. In this paper, tuples (*state, action, next state*) which indicate state transition are recorded into a table called state-transition table. Moreover, if an inconsistent state transition is probed, we regard such inconsistent state transition as contradiction and introduce two manners with the aim of resolving such contradiction.

3.1 State Classification from Perceptual Inputs by ART Neural Networks

In this paper, we adopt ART to realize the map from perceptual inputs to corresponding state. ART consists of two layers of neurons: F_1 and F_2 The neurons in the layers F_1 and F_2 are corresponding to a particular combination of sensory features and recognition code which represents states in the case of this paper, respectively. In the ART, given inputs are classified into the most resonant code that is decided by referring to a selection strength and vigilance criterion. If there are no resonant codes against certain inputs to classify, namely, there is no selection strength associated to code which is

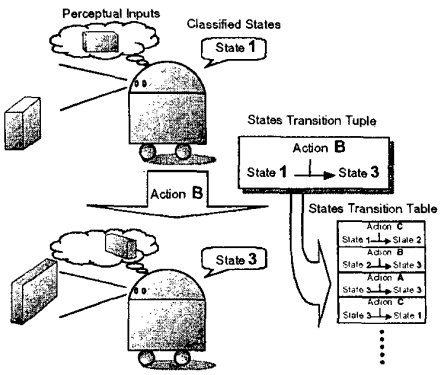

Fig. 1: A framework of the proposed method

greater than the vigilance parameter, a new recognition code is added to the level F_2 by adopting the input vector as the sample vector to the added recognition code.

Detailed description is shown in followings: Let $\mathbf{x}$ and $\mathbf{w}_i$ be a perceptual input vector for ART and sample vectors linked from all neurons in the level F_1 to a neuron i (recognition code) in the level F_2 . We adopt following selection strength T_i for state i (recognition code i) based on the distance between the input vector and the sample vector:

$$T_i = \frac{1}{\varepsilon_\alpha + \frac{|\mathbf{x} - \mathbf{w}_i|^2}{\varepsilon_\gamma + |\mathbf{x}|^2}}$$

where, ε_α and ε_γ indicate small positive constant values fixed in advance. For each state i, this selection strength T_i is calculated, and if the most resonant state i^*, i.e., a state which has the highest selection strength, is greater than the vigilance parameter σ, such state i^* is chosen as a state corresponding to the perceptual inputs. Otherwise, a new state j whose sample vector is the same as the perceptual inputs $\mathbf{x}$ is added into a set of sample vectors. That is,

$$\mathbf{w}_j = \mathbf{x}$$

Also, in the traditional ART, activated sample vector is updated for following to a current perceptual input vector described as follows:

$$\mathbf{w}_{i^*} = \beta\mathbf{x} + (1-\beta)\mathbf{w}_{i^*}$$

However, in the proposed method, state transition information is utilized vigorously so

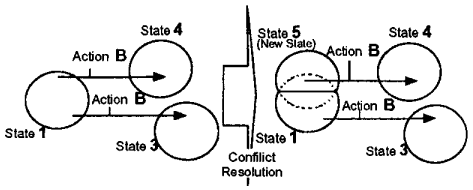

Fig. 2: A depiction of contradiction such that different results are caused by the same states and the same actions

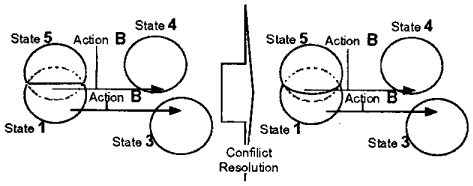

Fig. 3: A depiction of contradiction due to ambiguous states

that such improvement of an activated sample vector is not carried out.

3.2 Description of State Transition Table

In the proposed method, a state transition table is recorded in order to detect inconsistent transition. How to record the transition table is as follows: First, suppose that a state **1** corresponding to a perceptual input $\mathbf{x}_1$ is classified by the means of the previous section at a current time step. Moreover, suppose that, at that time step, the agent behaves an action **B** and, as a consequence of the action, next perceptual input $\mathbf{x}_2$ is received and classified into a state **2** at the next time step. Such process is recorded that

$$f_{\mathbf{B}}(\mathbf{1}) = \mathbf{2}$$

Note that state transitions with respect to ambiguous states which mean that selection strengths for several states exceed the vigilance criterion is not recorded into the state transition table.

3.3 Contradiction Resolution Mechanism

In this paper, we adopt two kinds of the notions of contradiction to constitute states: "different results caused by the same states and the same actions" and "contra-diction due to ambiguous states." Following subsections introduce them.

Different Results Caused by the Same States and the Same Actions

Suppose that there is a record in the state transition table such that an action **B** in a state **1** brought out a state **3**,

$$f_{\mathbf{B}}(\mathbf{1}) = \mathbf{3}$$

Moreover, now, the same action **B** in the same state **1** causes the different state **4**.

$$f_{\mathbf{B}}(\mathbf{1}) = \mathbf{4}$$

Above equations are inconsistent each other. In this case, it is possible that inadequate mapping from perceptual inputs to states is carried out by ART. Therefore, a new sample vector $\mathbf{w}_5$ which indicates a new state **5** is added to the ART by using a perceptual input $\mathbf{x}_5$ which causes above contradiction, i.e., $\mathbf{w}_5 = \mathbf{x}_5$. That is,

$$f_{\mathbf{B}}(\mathbf{5}) = \mathbf{4}$$

There is no contradiction in a state transition table as depicted in Fig. 2. Because the addition of the new state **5** affects to not only the definition of the state **1** but also neighbor states around the new state **5**, all records in the state transition table are destroyed at the time step. By adopting such destruction of the state transition table, meaningless detections of further contradiction are prevented.

Contradiction due to Ambiguous States

In the case of that several states are resonant simultaneously, called ambiguous states in this paper, even if the contradiction in the sense of last subsection is occurring, other resonant state might be consistent with the transition table. If such consistent resonant state is found, a new state by the means of the former subsection is not generated. Instead, following process is carried out: In the ambiguous state, instead of selection strength T_i described in section 2.1, biased selection strength T_i' for state i is used to decide a state for certain perceptual input.

$$T_i' = T_i + \sum_{j \in Ambiguous(i)} b(i, j)$$

where $Ambiguous(i)$ and $b(i, j)$ denote states which are in ambiguous state with state i for current perception and a bias term of state i against j, respectively. $b(i, j)$ is positive value and is updated as follows: if state i is consistent for state transition table,

$$\Delta b(i, j) = \delta$$

where δ is constant value fixed in advance. By introducing this mechanism, an excess of state generation is avoided.

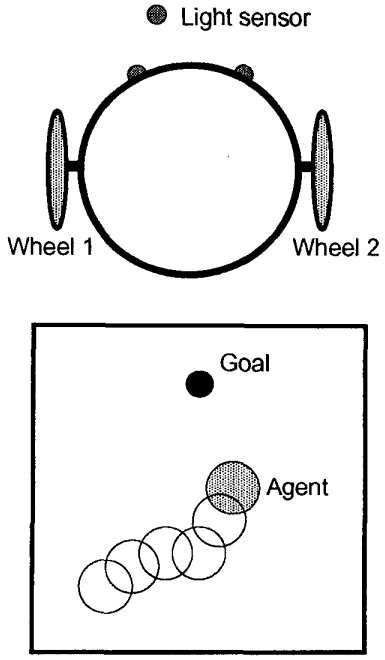

Fig. 4: An illustration of robot navigation problems

4 Computational Simulation

4.1 Experimental Environments

This paper examines the proposed method on robot navigation problems. Fig. 4 illustrates the outline of a mobile robot used in the experimental environment. The small mobile robot has two light sensors located in the front of the mobile robot. The outputs by the light sensors vary from 50 to 500 in accordance with the light intensity. The outputs of the light sensors decrease as the mobile robots approaches to the light source. Besides, actions of the agent consist of "go straight," "turn left," and "turn right." The mobile robot, i.e., the agent, has to learn perception-action rules in order to reach the light source smoothly. An episode consists of number of steps, i.e., tuples of perception and action, until agent reaches the light source or agent collides to the wall. At the beginning of each episode, initial direction and position of the agent are randomly determined. The criterion whether the agent can reach to the light source is set such that the distance between the agent and the light source is shorter than predefined length.

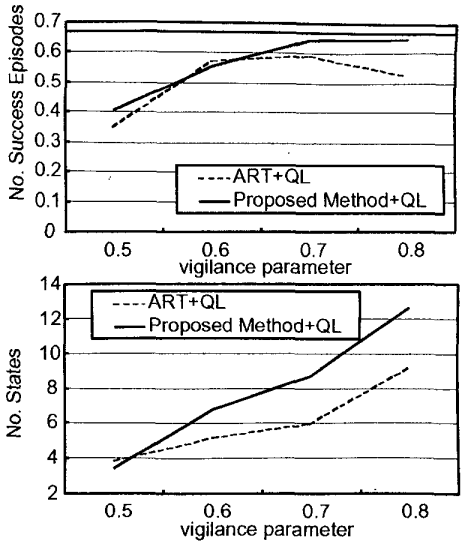

Fig. 5: Success ratio (UPPER) and number of segmented states (LOWER)

In this paper, two kinds of algorithms are examined in order to evaluate the effectiveness of the proposed method:

A: State construction method by ART with QL

B: Proposed state construction method with QL

4.2 Simulation Results

The vigilance parameter of ART is set to be one of 0.5, 0.6, 0.7, and 0.8. The learning parameters for Q-Learning, i.e., α and γ, are set to be 0.1 and 0.1, respectively. The reward given if the agent can reach the light source and the penalty punished if the agent collides to the wall are defined as +100 and −10, respectively. Both of state construction method and Q-learning are carried out until the number of episodes becomes 100. Succeeding 50 episodes when our proposed method is not carried out for state construction and Q-learning are used to investigate the number of success episodes. The proportion of reaching to goal is shown in Fig. 5. Plotted data denotes the averaged results over 100 experiments for each vigilance parameter.

Moreover, Fig. 6 depicts constituted state space in the case of that the vigilance parameter = 0.5 or 0.8. x axis and y axis denote the output by right-hand sensor and left-hand sensor, respectively. As delineated in this figure, the method **A** cannot yield proper state constitution at far area from the light source. The agent learns

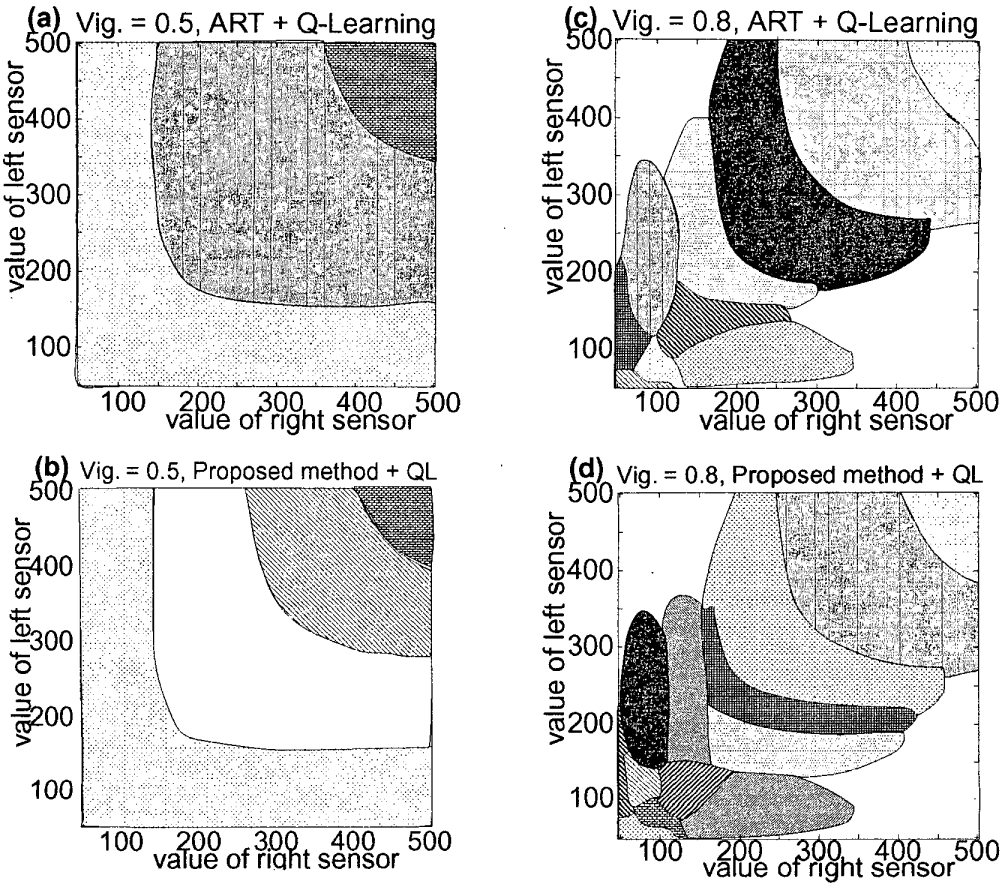

Fig. 6: State-space constructed by ART with Q-Learning (ART+QL) and the proposed method + Q-learning; (a) ART+QL with Vigilance parameter = 0.5, (b) the proposed method with Vig. = 0.5, (c) ART+QL with Vig. = 0.8, and (d) the proposed method with Vig. = 0.8

turn left or right at such area in order to avoid collisions to the wall. If the agent acquire rough state constitution such like Fig.6(a), the agent continues to turn right or left permanently. That is, the agent fails into "perceptual alias problem" in such area. That is the reason why the method **A** doesn't work well when the vigilance parameter is set to be lower value as depicted in Fig. 5. Both methods exhibit good performance with higher vigilance parameters. Such higher vigilance parameters are quite adequate for the robot navigation problem examined in this paper.

5 Conclusion

This paper applied our adaptive state construction method by ART Neural Networks to the robot navigation problems. However we adopted Q-Learning as reinforcement learning algorithms, the proposed method performs well with other kinds of reinforcement learning algorithms which can treat discrete states and actions because the proposed pays attention to just perception-action sequences. The nature of state-constitution acquired by the proposed method does not depend on the nature of reinforcement learning algorithms used with the proposed method but depends on the latent structure of tasks. As delineated in Fig. 6, it keeps the rationality for tasks in acquired state-constitution.

The contradiction resolution mechanism introduced in the proposed method serves as a mean of reducing the affection of lower vigilance parameter in comparison with ART+ QL. It causes easily achievement of tasks. However, the number of states is generally increased so that

the learning speed with the proposed method decreases for higher vigilance parameters.

Finally, we conclude the guideline of how to utilize the proposed method effectively as follows: first, the value of the vigilance parameter is set to be lower one. Then, the value increases until acquiring adequate number of states. In this case, the contradiction resolution mechanism works quite well.

As future work, we will incorporate a merge mechanism into the proposed method in order to acquire more proper state constitution autonomously.

Reference

1. H. Handa, A. Ninomiya, T. Horiuchi, T. Konishi and M. Baba, An Incremental State-Segmantation Method for Reinforcement Learning Using ART Neural Network, Proceedings of the 2000 IEEE International Conference on Industrial Electronics, Control and Instrumentation, pp.2732-2737, 2000.
2. R. Sutton and A. Barto, Reinforcement Learning, The MIT Press, 1999.
3. C. Watkins and P. Dayan, Q-learning, Machine Learning, Vol.8, pp.279-292, 1992.
4. R. Sutton, Learning to Predict by the Method of Temporal Differences, Machine Learning, Vol.3, pp.9-44, 1988.
5. A. Dubrawski and P. Reignier, Learning to Categorize Perceptual Space of a Mobile Robot Using Fuzzy-ART Neural Network, Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems IROS'94, Vol.2, pp. 1272-1277, 1994.
6. H. Murao and S. Kitamura, Incremental State Acquisition for Q-Learning by Adaptive Gaussian Soft-max Neural Network", Proceedings of the 1998 IEEE ISIC/CIRA/ISAS Joint Conference, Gaithersburg, pp.465-470, 1998.
7. H. Murao and S. Kitamura, Incremental Quantization of the Continuous Sensor Space for Leaning Agents, Intelligent Autonomous Systems, Y. Kakazu *et. al.*(Eds.), IOS Press, pp.272-279, 1998.
8. K. Yamada, M. Svinin, K. Ueda, Reinforcement Learning with Autonomous State Space Construction using Unsupervised Clustering Method, Proceedings of the 5th International Symposium on Artificial Life and Robotics, pp.850-853, 2000.
9. L. J. Lin, Self-Improving Reactive Agents: Case Studied of Reinforcement Learning Frameworks, From Animals to Animats, The MIT Press, pp.297-305, 1992.
10. L. Chrisman, Reinforcement Learning with Perceptual Aliasing, Proceedings of AAAI-92, pp.183-188, 1992
11. G. Drescher, MADE UP MINDS, The MIT Press, 1991.
12. M. Snorrason and A. Caglayan, Generalized ART2 Algorithms, World Congress on Neural Networks, 1994.